[Patent Document]

Japanese Patent Laid Open
No. 2003-303055

Disk storage system having disk arrays connected with disk adaptors through switches

Hitachi, Ltd.

Inventor(s):Tanaka, Katsuya ;Fujimoto, Kazuhisa

Application No. 10/212882, Filed 20020807, A1 Published 20031009

Abstract:

A disk storage system has high throughput between a disk adapter of a disk controller and a disk array. The disk adapter of the disk controller is connected to the disk array through switches. Data on a channel between the switch and a RAID group is multiplexed in the switch to be transferred onto a channel between the switch and the disk adapter and data on the channel between the switch and the disk adapter is demultiplexed in the switch to be transferred onto the channel between the switch and the RAID group. A data transfer rate on the channel between the disk adapter and the switch is made higher than that on the channel.

US.Class: 711114 711154

## ディスクアダプタとディスクアレイをスイッチをかして接続したディスク装置

S

特開2003-30305

€ 翐 ধ 罪 华 噩 (12) **(2**) (19) 日本国の日か (1 b)

**特開2003-303055** (P2003-303055A) (11)条件出限公园每号

			(43)公開日	(43)公開日 平成16年10月24日(2003,10.24)	3 (2003, 10, 24)
Ş	क्रिसास क	P.I		1	F-73-1-(\$2\$)
95/n.	301	G 0 6 F	3/06	3 0 1 M	5 B 0 6 5
				301B	
	7 O F			302A	
	540			540	

		大型片型	数世間米 米諾米 間米虫の数8 〇L (全14 月)
(21) 田岡毎年	(VER2002 106262(P2002 106262)	(71) 出國人	(71) 出國人 000005108
日間(22)	平成14年4月9日(2002.4.9)		株式会社日立與作所 東京都千代田区神田鑿柯台四丁目 6
		(72) 発明体	田中、脚也
			京京都国分号市京宏ケ盛一丁目280番地
			株式会社日立政作所中央研究所内
		(72) 発明者	摩本 和久
			東京都国分學市東藍ケ盟一丁目280番地
			株式会社日立段作所中央研究所内
		(74)代理人	100099298
			井西士 伊羅 俗 (外1名)
		トターム(数	ドターム(数号) SBOGS BADI CADA CADT CA12 CA15
-			CA30 CC08 CE11 CH01

(54) [発明の名称] ディスクアグプタとディスクアレイをスイッチを介して接続したディスク報章

(57) (現地)

[欧図] ディスクコントローラのディスクアダプタと ディスクアレイ間のスルーブットが高いディスク毎回を 位供することにある。

P ダブタ (DKA) とディスクアレイ (DA) をスイッチ (SW1, SW スイッチ (STI) とディスクアダブタ (DKA) 間のチャネル (D 01) 上のデータをスイッチ (SW1) において迎多瓜化してス ータをスイッチ (S¶1)において多皿化してスイッチ (S¶1) 【解袂年段】 ディスクコントローラ(DXC)のディスク Dグループ(81)間のチャネル(011,012,013,014)上のデ とディスクアダプタ (DKA) 閉のチャネル (DO1) に転送し、 イッチ (STI) とRAIDグループ (RI) 間のチャネル (DII, 2, SW3, SW4)を介して接続する。スイッチ(SW1)とRA J 012, 013, 014)に低送する。ディスクアダプタ (DKA) とス イッチ (S#1)間のチャネル (D01) 上のデータ伝送選度を、 チャネル(DI1, DI2, DI3, DI4)のデータ低送遊度より高く

O

が接続された各ポートとの間でのポート間の接続の切り 前配ディスクアダプタと前配ディスクアレイを、パッフ 前配スイッチは、前配ディスクアダプタが接続されたポ アレイからなり、前配ディスクコントローラはチャネル からなり、前紀ディスクコントローラはチャネルアダプ **ートと前記ディスクアレイを構成するディスクドライブ** 換えを、入力されたフレーム年に、故フレーム内の送信 ディスクコントローラとディスクアレイ タとキャッシュメモリとディスクアダプタを有するディ ディスクコントローラと複数のディスク **先情報にしたがって行うことを特徴とするディスク藝** アメモリを有するスイッチを介して接続し、 スク装置において、 [開水項2] 年度15年10月24日(2002)10 24)

前記ディスクアレイはループ状に接続した複数のディス アダプタとキャッシュメモリとディスクアダプタを有す クドライブからなり、前紀ディスクアダプタと前紀複数 のディスクアレイとをパッファメモリを有するスイッチ るディスク装置において、

前配ディスクアダプタと前配スイッチ間のチャネル当り データ転送速度を、前配スイッチと前配複数のディスク アレイ間のチャネル当りデータ転送速度より高く設定

前記スイッチは、前紀ディスクアダプタが接続されたポ ートと前配複数のディスクアレイが接続された各ポート との間でのボート間の接続の切り換えを、入力されたフ レーム毎に、散フレーム内の送信先情報にしたがって行 うことを特徴とするディスク装置。

「群求項3】 ディスクコントローラとディスクアレイ: からなり、前起ディスクコントローラはチャネルアダプ タとキャッシュメモリとディスクアダプタを有するディ スク装置において、

前配ディスクアダプタと前配ディスクアレイを、パッフ アメモリを有するスイッチを介して接続し、

可一のスイッチに接続したディスクドライブの組み合わ 前紀ディスクアダプタと前紀スイッチ間のチャネル当り せでRAIDグループを構成し、

データ転送速度を、前記スイッチと前記ディスクアレイ

前配スイッチは、前配ディスクアダプタが接続されたボ り換えを、入力されたフレーム毎に、放フレーム内の送 間先情報にしたがって行うことを特徴とするディスク藝 プが接続された各ポートとの間でのポート間の接続の切 ートと前紀RAIDグループを構成するディスクドライ 間のチャネル当りデータ転送速度より高く設定し、

**第1のディスクコントローラは第1のチャネルアダプタ** 【開求項4】 第1のディスクコントローラと第2のデ イスクコントローラと複数のディスクアレイからなり、

## ディスクアダプタとディスクアレイをスイッチをかして接続したディスク装置

(や酢醇状の範囲)

S

S

0 က - 30

特開2003

と第1のキャッシュメモリと第1のディスクアダプタモ

と第2のキャッシュメモリと第2のディスクアダプタを 第2のディスクコントローラは第2のチャネルアダプタ 有するディスク装置において、

第1のディスクアダプタと前配複数のディスクアレイと をパッファメモリを有する第1のスイッチを介して接続 し、且つ第2のディスクアダプタと前配複数のディスク アレイとをパッファメモリを有する第2のスイッチを介 ダプタを接続し、第2のスイッチと第1のディスクアダ 10 して接続し、さらに第1のスイッチと第2のディスクア

第1のディスクアダプタと第1のスイッチ間、および算 りデータ転送速度を第1のスイッチと前配複数のディス クアレイ団のチャネル当りデータ転送選段より高く設定 2のディスクアダプタと第1のスイッチ間のチャネル当 ブタを接続し、

第2のディスクアダプタと第2のスイッチ間、および類 1のディスクアダプタと第2のスイッチ間のチャネル当 りデータ転送速度を第2のスイッチと前配複数のディス クアレイ間のチャネル当りデータ気送強度より高く設定 2

第1のスイッチは、第1のディスクアダプタまたは第2 15 ィスクアレイが核焼された各ポートとの間でのポート間 の接続の切り換えを、入力されたフレーム年に、鼓フレ のディスクアダプタが接続されたポートと前配複数のデ **一ム内の送信先情報にしたがって行い、** 

第2のスイッチは、第1のディスクアダプタまたは第2 のディスクアダプタが接続されたボートと前配複数のデ 4 スクアレイが接続された各ポートとの凹でのポート回 の接続の切り換えを、入力されたフレーム年に、眩フレ **--ム内の送信先情報にしたがって行うことを特徴とする** ディスク装団。 8

「酵求項5】 第1のディスクコントローラと第2のデ 育1のディスクコントローラは第1のチャネルアダプタ と第1のキャッシュメモリと第1のディスクアダプタを 35 イスクコントローラと複数のディスクアレイからなり、

と第2のキャッシュメモリと第2のディスクアダプタを 第2のディスクコントローラは第2のチャネルアダプタ 有するディスク数囮において、 ÷

をパッファメモリを有する第1のスイッチを介して抜税 アレイとをパッファメモリを有する第2のスイッチを介 第1のディスクアダプタと前配複数のディスクアレイと し、且つ第2のディスクアダプタと前配複数のディスク ダブタを接続し、第2のスイッチと第1のディスクアダ して接続し、さらに第1のスイッチと第2のディスクア ş

第1のディスクアダプタと第1のスイッチ間、および第 2のディスクアダプタと第1のスイッチ間のチャネル当 プタを接続し、 20

-

ディスクアダブタとディスクアレイをスイッチ を介して接続したディスク装置

りデータ 低送速度を算1のスイッチと前配複数のディス クアレイ間のチャネル当りデータ転送遊取より高く設定

1のディスクアダプタと第2のスイッチ間のチャネル当 クアレイ団のチャネル当りデータ転送速度より高く設定 第2のディスクアダプタと第2のスイッチ間、および第 りデータ転送速度を算2のスイッチと前配複数のディス

ダプタと辞2のスイッチ間を被殺したチャネルと同等の ダブタと 抑1のスイッチ間を 協校したチャネルと同等の のディスクアダプタ虫たは第2のスイッチが接続された 第1のスイッチと類2のスイッチを、類1のディスクア ゲータ転送遠度を有するチャネルと、 類2のディスクア 第1のスイッチは、第1のディスクアダプタまたは第2 ポートと前記複数のディスクアレイが接続された各ポー トとの間でのポート間の接続の切り換えを、入力された フレーム毎に、酸フレーム内の送信先情報にしたがって データ転送速度を有するチャネルと、を介して接続し、

第2のスイッチは、第1のディスクアダプタまたは第2 トとの間でのボート間の接続の切り換えを、入力された フレーム毎に、抜フレーム内の送信先情報にしたがって のディスクアダプタまたは第1のスイッチが接続された ポートと前配複数のディスクアレイが接続された各ポー 行うことをわ数とするディスク模団

[酵求項6] 開求項1乃至間求項5のいずれかの酌求 信配ディスクアレイからのデータ脱み出し時には、信配 前配スイッチにおいて多皿化して前配ディスクアダプタ ディスクアレイから付記スイッチに転送されるデータを **項記載のディスク製置において、** 

値配スイッチにおいて逆多瓜化して値配ディスクアレイ イスクアダプタから仰配スイッチに伝送されるデータを 何配ディスクアレイへのデータ母き込み時には、何配デ に転送することを仲徴とするディスク模型。

[請求項7] 請求項1乃至額求項5のいずれかの額求 ディスクアダブタからディスクアレイへのデータ哲食込 み時に、前紀ディスクアダプタは、前記ポート間の接続 **頃記載のディスク模型において、** 

の切り替えが周期的に行われるように、送出するフレー ディスクアレイからディスクアダプタへのデータ観み出 **り時に、色的スイッチは、 かケンドロガン方式により**哲 紀ポート間の抜放を切り替えることを特徴とするディス ムに送伯先信頼を設定し、

る比、と同程度に設定することを仲徴とするディスク藝 イッチ間のチャネル当りデータ転送꿟取の、スイッチと えるボート数を、ディスクアダプタとス ディスクアレイ間のチャネル当りデータ転送遊度に対す **開収項7記載のディスク装置において、** 国歴色にむり [原水斑8]

間求項1乃至請求項5のいずれかの請求 項記載のディスク装置において、 (間水田)

ーブルで接続し、前記スイッチと前記ディスクアレイ関 をメタルケーブルで接続することを特徴とするディスク **帕妃ディスクアダプタと前配スイッチ間を光ファイパケ** 

[発明の詳細な説明]

(0001)

テムにおける2次記憶装置に関し、特に入出カデータ転 [発明の属する技術分野] 本発明は、コンピュータシス 送性能が高いディスク装置に関する、

配徴装置に保存され、CPUなどが必要とするときに応 じて2次配位装置に対してデータの哲き込みおよび脱み 高度情報化に伴い、コンピュータシステムにおいて、こ は、CPU(中央処理装配)が必要とするデータは2次 出しを行う。この2次記憶装置としては、一般に不揮発 な記憶媒体が使用され、代表的なものとして磁気ディス ク装団や、光ディスクなどのディスク装配がある。近年 【従来の技術】現在のコンピュータシステムにおいて の種の2次記憶装置の高性館化が要求されている。 20

[0003] 図9に、従来のディスク装置のブロック図 を示す。図9において、ディスク装置はディスクコント ローラDKCとディスクアレイDAで構成される。ディ スクコントローラDKCは、上位側CPU(図示せず)

ディスクアレイDAに対して脱みむきするデータを一時 保存するキャッシュメモリCMと、ディスクコントロー とディスク藝団を接続するチャネルアダプタCHAと、

HAはC1、C2、C3、C4の4本のチャネルでCP ラDKCとディスクアレイDAを接続するディスクアダ シュメモリCMとディスクアダプタDKAは、バスまた はスイッチで相互接続されている。チャネルアダプタC Uと接続している。ディスクアダプタDKAはD1、D 2. D3. D4の4本のチャネルでディスクアレイと接 プタDKAからなる。チャネルアダプタCHAとキャッ 睨している。ここでディスクアレイDAはディスクグル -7R1. R2. R3. R4からなる。ディスクアレイ

1. R2. R3. R4がそれぞれRAIDグループを構 DAにおいてRAIDシステムを構築する場合は、R

[0004] チャネルC1、C2、C3、C4から入力

ヤネルにはプロック単位に分割されたデータを、残りの 内に数当データの有無を聞べる。有る場合は、キャッシ された母き込みデータは、キャッシュメモリCMに放デ 一クを母き込むと同時に、 放データをプロックサイズ単 位に分割し、チャネルロ1、ロ2、ロ3、ロ4の内3チ ディスクアダプタDKAからディスクアレイDAへ送信 する。データ競み出し時は、先ずキャッシュメモリCM 1 チャネルは前記分割データから計算したパリティを、 S

ッシュメモリ内幌み出しデータをCP Uへ送信する。キ ヤッシュメモリCM内に無い場合にディスクアダプタD KAは、D1、D2、D3、D4を介してディスクアレ し、チャネルアダプタCHAを介して飲み出しデータを CPUへ送信する。 この種の従来技術を第1の従来技 術と呼ぶ。第1の従来技術として関連するディスク装置 インフレーム, 981 (1998年) 第144頁から第 ュメモリCMからチャネルアダプタCHAを介してキャ は、例えば、日経BP社刊の「日経コンピュータ別師メ イDAからプロック単位に分割されたデータを脱み出 153頁に記載されているディスク装置がある。

[0005] ディスクアダプタとディスクアレイを、ス 4の交換デバイス」に開示されている。以下、敵公報に 記載の従来の技術を第2の従来技術と呼ぶ。第2の従来 の技術によれば、ディスクアレイに関連したパス本教と **制御ブロックを介して接続したディスク装置が、特開平** 6-19627号の「回転形配像装置」に開示されてい イッチを介して接続したディスク装団が、特関平5-1 73722号の「マルチチャンネルデータおよびパリテ ディスクアダプタに関連したパス本数とを独立に散定で る。以下、眩公報に配載の従来の技術を第3の従来技術 と呼ぶ。類3の従来技術によれば、ディスクアダプタと きる。ディスクアダプタとディスクアレイを、バッファ ディスクアレイ間のデータ転送速度を任意に設定でき、 ディスクの回転待ちの影響を低減できる。

ネルでは、現状でチャネル当りのデータ転送速度が1G エンドに比く高速化が購しいこと、である。第2の理由 は、たとえディスクドライブが高速化したとしても、全 歩に伴い、1チャネル当りのデータ転送速度は年々増加 している。例えばディスク芸型に使われるファイバチャ bpsから2Gbpsであるが、近い将来4Gbpsか CPUとチャネルアダプタ間(以下フロントエンドと呼 ぶ)のスールプットはこの高速化に従うことが予想され れる。第1の理由は、ディスクドライブは機械部品を含 てのディスクドライブ毎に高速インターフェイスを搭載 するのは、多数のディスクドライブを有するディスク装 [発明が解決しようとする課題] ネットワーク技術の造 由により、フロントエンドほど高強化されないと予想さ むので、電子、光楽子のみ高速化を行えば良いフロント (以下パックエンドと呼ぶ) のスループットは以下の理 510Gbpsへ萬滋化されることが予定されている。 る。ところが、ディスクアダプタとディスクアレイ間 聞の高コスト化を招くことである。

[0007] 第1の従来技術では、チャネルアダプタの トを多数、ディスクアダプタに散けることも考えられる チャネル当りのデータ転送速度を向上させても、フロン トエンドとパックエンドのスループット転離により、デ ィスク装置の性能が向上しないという問題があった。ま た、パックエンドのスループット向上のために低速ポー

ディスクアダプタのポート数増加は何仰を複雑とす る。 第2の従来技術では、ディスクアダプタとディスク アレイとの関にスイッチを適用することによりディスク 増設ポート数を増加させることができるが、チャネル当 りのデータ転送速度はディスクアレイのデータ転送速度 に朗吸されるので、ディスクアダプタとディスクアレイ 聞のスループットが性能ネックになるという周園があっ た。 第3の従来技術は、ディスクの回転特ち時回の影響 を低級できる技術であり、フロントエンドとバックエン ドのスループット範疇は低減できないという問題があっ

スク装置を提供することにある。本発明のさらに他の目 タとディスクアレイ間ネットワークを有するディスク数 [0008] 本発明の目的は、ディスクアダプタとディ スクアレイ間のスループットが高いディスク数回を提供 することにある。本発明の他の目的は、ディスクアダプ 15 タとディスクアレイقのスループットが描く、 且つディ スクドライブ接載台数が多いディスク装置を提供するこ とにある。本発明のさらに他の目的は、信頼性が高いデ 4.スクアレイを有するディスク藝囮を提供することであ る。本発明のさらに他の目的は、信頼性が高いディスク アダプタとディスクアレイ間ネットワークを有するディ 的は、偕飯性およびスループットが高いディスクアダブ ディスクからの破み出しおよびディスクへの仰きこみを **高スループット化できるディスク数団を提供することに** ある。本発明のさらに他の目的は、高スループットを維 **怜できるディスク装団を提供することである。本発明の 配を提供することにある。本発明のさらに他の目的は、** 

さらに他の目的は、高スループットで低コストなディス ク装置を提供することである。 22

め、本発明は、ディスクコントローラとディスクアレイ 【喋盟を解決するための手段】上配目的を遊成するた [6000]

スクドライブが接載された各ポートとの間でのポート間 からなり、ディスクコントローラはチャネルアダプタと キャッシュメモリとディスクアダプタを有するディスク 抜置であり、ディスクアダプタとディスクアレイを、パ ッファメモリを有するスイッチを介して接続し、ディス クアダプタとスイッチ回のチャネル当りデータ転送遊取 を、スイッチとディスクアレイ間のチャネル当りデータ **院送遊段より笛く設定し、スイッチは、ディスクアダブ** の接続の切り換えを、入力されたフレーム年に、放フレ **一ム内の送信先情報にしたがって行っている。また、前** 紀ディスクアレイはループ状に依頼した複数のディスク ドライブからなり、疝配ディスクアダプタと疝配複数の ディスクアレイとをパッファメモリを有するスイッチを 介して接続し、ディスクアダプタとスイッチ間のチャネ タが後載されたポートとディスクアレイを構成するディ ル当りデータ転送遊度を、スイッチと複数のディスクア \$

2

アイ阿のチャネル辿りゲータ院说協協より超へ設成し、

4%

た、他紀ディスクアダプタと仰紀ディスクアレイを、パ ッファメモリを有するスイッチを介して接板し、同一の スイッチは、ディスクアダプタが接板されたポートと複 数のディスクアレイが接続された各ポートとの間でのポ スイッチに接続したディスクドライブの組み合わせでR AIDグループを構成し、ディスクアダプタとスイッチ 間のチャネル当りデータ転送速度を、スイッチとディス クアレイ間のチャネル当りデータ転送速度より高く設定 女フレーム内の送信先信報にしたがって行っている。 虫 し、スイッチは、ディスクアダプタが接続されたポート とRAIDグループを傾成するディスクドライブが接続 **一ト間の後戌の切り換えを、入力されたフレーム低に、** された各ポートとの間でのポート間の接線の切り換え

報にしたがって行っている。また、第1のディスクコン のチャネルアダプタと第1のキャッシュメモリと第1の は筑2のチャネルアダプタと第2のキャッシュメモリと を、入力されたフレーム低に、散フレーム内の送信先情 トローラと抑2のディスクコントローラと複数のディス クアレイからなり、抑1のディスクコントローラは第1 ディスクアダブタを有し、第2のディスクコントローラ 訊2のディスクアダプタを有するディスク藝匠であり、

間のチャネル当りデータ転送速度を第2のスイッチと前 第1のディスクアダプタと前配複数のディスクアレイと アレイとをパッファメモリを有する第2のスイッチを介 をパッファメモリを有する第1のスイッチを介して接続 し、且つ第2のディスクアダプタと前記複数のディスク して接続し、さらに第1のスイッチと第2のディスクア ダブタを接続し、第2のスイッチと第1のディスクアダ プタを接続し、第2のディスクアダプタと第2のスイッ チ間、および第1のディスクアダプタと第2のスイッチ 記複数のディスクアレイ間のチャネル当りデータ転送選 敗より商く設定し、第1のスイッチは、第1のディスク

アダプタまたは第2のディスクアダプタが接続されたポ との間でのポート間の接続の切り換えを、入力されたフ フーム母に、数ファーム内の辺徊先疳殻につたがって行 い、第2のスイッチは、第1のディスクアダプタ虫たは 第2のディスクアダプタが接破されたポートと前配複数 ト間の接続の切り換えを、入力されたフレーム毎に、厳 と、第2のディスクアダプタと第1のスイッチ間を接続 **ートと印記複数のディスクアレイが接続された各ポート** 上配路1のディスクアダプタと第2のスイッチ間を接続 したチャネルと同等のデータ転送速度を有するチャネル からのデータ競み出し時には、何配ディスクアレイから のディスクアレイが依依された各ポートとの間でのポー したチャネルと同等のデータ転送遊成を有するチャネル た、さらに、上配算1のスイッチと第2のスイッチを、 と、を介して投税している。また、前配ディスクアレイ フレーム内の送信先情報にしたがって行っている。虫

ッチにおいて逆多塩化して前配ディスクアレイに転送す るようにしている。また、前紀ディスクアダプタから前 記ディスクアレイへのデータ暫き込み時に、前配ディス クアダプタは、前記ポート間の接続の切り替えが行われ また、さらに、切り替えるポート数を、ディスクアダプ るように、送出するフレームに送信先情報を設定し、前 兄ディスクアレイから前配ディスクアダプタへのデータ 脱み出し時に、前記スイッチは、ラウンドロピン方式に スクアダプタと前記スイッチ間を光ファイパケーブルで スクアレイへのデータ音を込み時には、前記ディスクア ダブタから前配スイッチに転送されるデータを前配スイ タとスイッチ間のチャネル当りデータ転送速度の、スイ ッチとディスクアレイ間のチャネル当りデータ転送速度 に対する比、と同程度に散定している。また、前配ディ より前記ポート間の接続を切り替えるようにしている。

**【発明の奥施の形態】以下、図面を参照して本発明の実** 施の形態を詳細に説明する。図1に本発明の、第1の実 ケーブルで検焼するようにしている。 [0010]

**镣幌し、前記スイッチと前記ディスクアレイ間をメタル** 

施の形態であるディスク 挺四の構成を示す。 本実施の形 Cは、チャネルアダプタCHAと、キャッシュメモリC 娘のディスク装置は、ディスクコントローラDKCとデ **ィスクアレイDAからなる。ディスクコントローラDK** Mと、ディスクアダプタDKAからなる。チャネルアダ プタCHAは、上位CPU(図示せず)とディスクコン トローラDKCとがデータを送受信する際の制御を行

う。 C1、 C2、 C3 およびC4は、チャネルアダプタ CHAがCPUと通信するチャネルである。 キャッシュ メモリCMは、本実施の形態のディスク装置が入出力す るデータを一時保存するメモリである。 ディスクアダブ タDKAは、ディスクコントローラDKCとディスクア レイDAとがデータを送受値する際の街御を行う。ディ スクアダプタDKAは、チャネルD01、D02、D0 3、D04を介して、ディスクアレイDAと接続する。

ディスクアダプタDKAとディスクアレイDAを、スイ ッチSW1、SW2、SW3、SW4を介して接続して いる点に特徴がある。ディスクアレイDAは、ディスク DKAと接続する。同様に、ディスクグループR2はス イッチSW2を介して、ディスクグループR3はスイッ チSW3を介して、ディスクグループR4はスイッチS W4分して、それぞれディスクアダプタDKAと接続す グループR1, R2, R3, R4からなる。 ディスクグ ループR1は、スイッチSW1介してディスクアダプタ 【0011】ここで、本実施の形態のディスク装置は、

[0012] 本実施の形態のディスク装置においてRA 20

前配スイッチに転送されるデータを前配スイッチにおい て多皿化して何起ディスクアダプタに転送し、前配ディ

ディスクアダプタとディスクアレイをスイッチをかして接続したディスク装置

1, R2, R3, R4&, ENFARAID911-72 スクグループへのデータ航み出しまたは啓き込み時のデ **一夕からパリティが生成される。 放プロック単位に分割** する。本実施の形態では、4個のディスクドライブでR AIDグループを構成しているが、RAIDグループを 構成するドライブ数を4個に限るものではない。各ディ である。チャネルC1、C2、C3、C4からディスク グループR 1 へ番き込むためにCP Uから送信されたデ ータは、ディスクアダプタDKAにおいてブロック単位 に分割されると同時に、散ブロック単位に分割されたデ されたデータと、生成されたパリティは、チャネルD0 1を通りスイッチSW1へ入力される。スイッチSW1 は、RAID制御に伴い、敢プロック単位に分割された データと、生成されたパリティとをルーティングし、チ 1. D12. D13. D14を介してディスクグループ る. ここでR 1はRAIDレベル5のRAIDグループ IDシステムを構築する場合は、ディスクグループR ータの斑れを、ディスクグループR 1を例にして述べ R1からプロック単位に分割されたデータを飲み出し、 一夕競み出し時は、ディスクアダプタDKAは、D1 ヤネルD11, D12, D13, D14へ分配する。

[0013] 図9に示した従来のディスク装置では、デ スイッチSW1 通過後に別々のチャネルに分配される点 D3、D4上で、既にディスクアレイへの留き込みデー イスクアダプタDKAに接続したチャネルD1、D2、 タおよびパリティが別々のチャネルに分配されていた。 それに対し、本実施の形態のディスク装置においては、 して筋み出しデータを受信する。 が従来と異なる。

スイッチSW1でシリアル化して、チャネルD01を介

であるスイッチの動作を、スイッチSW1を例にとり説 [0014] 次に、本実施の形態のディスク藝匠の特徴 1、P2、P3、P4、P5は、全二直通信可能な入出 る。スイッチSW1の内部構成を図2と図3に示す。簡 明する。SW2~SW4の動作もSW1の動作と同様で カポートであり、ポート毎にパッファメモリを有してい D13、D14上を流れるデータは、フレーム単位で送 単のため、データの進行方向によりスイッチ動作を分け ある。図1に示すように、スイッチSW1は入出力ポー 受信され、かつデータは8B10B変数で符号化されて て説明する。また、チャネルD01、D11、D12、 トP1、P2、P3、P4、P5を有する。ポートP

ネルD01、D02、D03、D04上で全二重通信が

可能である。

ディスクアダプタDKAとディスクアレイDAは、チャ

5のクロスパスイッチであり、入力ポート1n1、in -- ムを入力し、ポートP2、P3、P4、P5から出力 する場合を示す。これはディスクアレイへの哲き込み時 [0015] 図2は、ポードP1からブロック内のフレ のスイッチ動作に相当する。スイッチSW1は図2に示 すように、クロスバスイッチXSWと、スイッチコント ローラCTLからなる。クロスパスイッチXSWは5×

out 2、out 3、out 4、out 5を有する。ポ **一トP1から入力したフレームは、シリアルバラレル変 換装置SP1と、パッファメモリBM1と、8B10B 変換デコーダDEC1を軽由し、スイッチコントローラ** て選ばれた場合は、入力したフレームは出力ポートou **CTLと入力ポート।n1へ入力される。スイッチコン** 掛かれた送信先アドレスを解脱し、クロスパスイッチX アメモリBM2と、パラレルシリアル変換装図PS2を 紙由して、ボートP2から出力される。ここで、パッフ 2. 1n3. in4. in5と、出力ボートout1, トローラCTLにおいて、入力フレームのヘッダ邸分に SWを切り換える。例として、ポートP2が出力先とし t 2 と、8 B 1 0 B変換エンコーダENC 2 と、パッフ

ラレルデータに敷換し、ポートP1におけるデータ転送 [0016]シリアルパラレル党徴装回SP1は、8B 10B符号化されたシリアルデータを10bit幅のパ 協度の1/10の強度に同期してパッファメモリBM1 First-out) メモリである。

アメモリBM1、BM2はFIFO (First-In

に巻き込む、8月10日デコーダDEC1は、クロスパ スイッチXSWの動作速度に同期して、10bltパラ レルデータをパッファメモリBM1から脱み出し、8B 5. 8B10BI>J-FENC2H, DUANAA" び8B10B符号化し、10b1tパラレルデータに変 チXSWでスイッチされた8bitパラレルデータを再 数後、クロスパスイッチXSWの動作速度に同切してパ ッファメモリBM2に毎き込む。パラレルシリアル変数 **疫団PS2は、ポートP2におけるデータ転送速度の1** /10の滋度に同期して、10bltパラレルデータを は、ポートPIにおけるデータ転送速度からボートP2 パッファメモリBM2から散み出し、シリアル化して、 10B質号化して、8bltパラレルデータに変換す ボートP2から出力する。以上によりスイッチSW1 25

と、ポートP2、P3、P4、P5から出力されるフレ 一ムを示した図である。故形の凸はフレームが存在する ある。図4では、入力ポートP1でのデータ低送遊度が 出力ポートP2、P3、P4、P5におけるデータ転送 速度のm倍あるとする。従って、ポートP1におけるフ る。フレームは伝送するデータ容量に従ってそのフレー ンシャルアクセスが行われており、フレーム母が一定で レームFb2の時間T1は、ボートP2からの出力時に ム長が変化するが、ここではディスクアレイへのシーケ [0017] 図4は、ポートP1へ入力するフレーム 時間、凹はフレームが存在していない時間を示してい におけるデータ転送速度へ速度変換する。 9

[00]18] 入力のデータ転送速度が避く、且つ出力の データ転送速度が遅い場合は、スイッチを周期的に切り 換えないと出力ポートのパッファメモリが強れ、スルー ブットが低下する。フレームがスループットの低下無く T3へ伸びている。 にこでT3=EXT1である。

20

ば、フレームの複数無く、スループットの角下は低この ない。T2213はn≥mと同じある。つまり、ディス クアレイへのデータ音を込み時に、スイッチにおいてス スイッチを迢迢するには、図4のように周辺的に出力ポ 一トを切り換える必要がある。スイッチ切替えポート数 をnとすると、スイッチ切替犬周加丁2~n×T1であ イッチ間のチャネル当りデータ転送速度の、スイッチと る(フレームの無い時間は無ねした)。 T2≥T3なら ルーブット低下を起こさないための条件は、周辺的に切 で周期的に切り替えることにより逆多皿化し、ポートP 2、P3、P4、P5へ分配して出力する。スイッチを **周期的に切り飲える方法の一つは、スイッチに接続した** り替えるスイッチポート数nを、ディスクアダプタとス ディスクアレイ間のチャネル当りデータ転送遊貨に対す る比田、以上に設定することである。この条件が保たれ れば、スイッチSW1は、ポートP1から入力したデー タをパッファメモリにおいて遊収変換し、フレーム単位 る。 RAIDのストライピング的仰に従えば、スイッチ ディスクグループをRAIDグループとすることであ は周期的に切り替わる。

す。これはディスクアレイからの脱み出し時のスイッチ [0019] 図3は、ポートP2、P3、P4、P5か 5フレームを入力し、ポートP1から出力する場合を示 動作に相当する。例えば、ポートP2から入力したフレ **一ムは、シリアルバラレル変換装置SP2と、バッファ** メモリBM2と、8B10B変換デコーダDEC2を経 由し、スイッチコントローラCTLと入力ポートin2 解脱し、クロスパスイッチXSWを切り換える。図3の 場合は、ラウンドロピン方式によりクロスパススイッチ 2. P3. P4. P5) に同時にフレームが個く。これ 入力フレームのヘッダ部分に僻かれた送信先アドレスを ら複数の入力フレームは同期して入力ポートに届く必要 はない。スイッチは、総当り的に入出力ポート間接税を フレームずつ出力ポート (P1) へ転汲する。このよう に、スイッチを総当り的に切り替える方式を、ラウンド により、結果的にスイッチは周期的に切り替わることに 切り替えることにより、これら複数の入力フレームを1 XSWを切り替えて、原命にポートP2、P3、P4、 ロビン(Round Robin)方式と呼ぶ。ラウンドロビン方式 へ入力される。スイッチコントローラCTLにおいて、 P 5から入力されるデータは全てポートP 1 へ出力す る。 すなわち、脱み出し時は、複数の入力ボート (P

10B符号化されたシリアルデータを10blt幅のパ 50 [0020]シリアルバラレル変換数図SP2は、8B

フレームは出力ポートのロ11と、8B10B変換エン コーダENC1と、バッファメモリBM1と、バラレル シリアル校校被回PS1を穏由して、ポートP1から出

ム内送信先信頼に従って切り替わることに適いはない。

なる。なお、親み出し時においても、スイッチはフレー

ラレルデータに変換し、ポートP2におけるデータ転送 にむき込む, 8B10BデコーダDEC2は、クロスパ 速度の1/10の速度に阿姆してパッファメモリBM2 スイッチXSWの動作速度に同期して、10bit/パラ レルデータをパッファメモリBM2から杭み出し、8B 5. 8B10BI>J-YENCIH, DDAKA チX SWでスイッチされた8bitパラレルデータを再 10日復号化して、8biiパラレルデータに変換す

び8B10B符号化し、10bitパラレルデータに変 校後、クロスパスイッチXSWの動作選度に同期してパ は、ポートP2におけるデータ転送適度からポートP1 ッファメモリBM1に毎き込む。パラレルシリアル変数 装置PS1は、ポートP1におけるデータ転送速度の1 /10の選段に同期して、10ち1tパラレルデータを パッファメモリBM1から睨み出し、シリアル化して、 ポートP1から出力する。以上によりスイッチSW1 におけるデータ転送遠度へ遠度変換する。

[0021] 図5は、ポートP2、P3、P4、P5へ ムを示した図である。彼形の凸はフレームが存在する時 が政化するが、ここではディスクアレイへのシーケンシ る。図5では、入力ポートP1でのデータ転送速度が出 カポートP2、P3、P4、P5におけるデータ転送速 度のm倍あるとする。従って、ポートP5におけるフレ 入力するフレームと、ポートP 1 から出力されるフレー フレームは伝送するデータ容量に従ってそのフレーム長 間、凹はフレームが存在していない時間を示している。 ャルアクセスが行われており、フレーム長が一定であ

ムFe2、Fe3、Fe4、Fe5をポートP1から出 **一ムFe5の時間T4は、ポートP1からの出力時にT** 力するのにかかる時間をT6とする。 スイッチ切り替え ポート数をnとすると、T6≒n×T5である (フレー ムの無い時間は無視した)。 スイッチにおいて幅像によ るスループット低下を防止するためには、T6≤T4と 5へ縮んでいる。ここでT4=m×T5である。フレー する必要がある。T6≤T4はn≤mと同じある。

[0022] つまり、ディスクアレイからのデータ脱み 出し時に、スイッチにおいてスループット低下を起こさ ないための条件は、周期的に切り替えるスイッチポート 数nを、ディスクアダプタとスイッチ間のチャネル当り データ転送速度の、スイッチとディスクアレイ間のチャ タをパッファメモリにおいて速度変換し、フレーム単位 で周期的に切り替えることにより多盟化し、ポートP1 ト数を、ディスクアダプタとスイッチ間のチャネル当り データ転送速度の、スイッチとディスクアレイ間のチャ る、ことである。この条件が保たれれば、スイッチSW 1は、ポートP2、P3、P4、P5から入力したデー へ出力する。よって、ディスクアレイへの警ぎ込みおよ びディスクアレイからの競み出しを高スルーブット化す るためには、n 与m、つまり、周期的に切り えるボー ネル当りデータ転送速度に対する比m、以下に設定す

ネル当りデータ転送速度に対する比、と同程度に設定す

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

ì

i

S

特別2003-30305

ので、実効的なスルーブットは2Gbps×4=8Gb [0023] 例えば、ディスクアダプタとスイッチ間の 4Gbpsのチャネル1本で接続し、スイッチとディス クアレイ間を16ちpsのチャネル4本で接続する。ま た.. ディスクアダプタとスイッチ間の10Gbpsのチ ャネル1本で接続し、スイッチとディスクアレイ間を2 Gbpsのチャネル4本で接続する。この場合、スイッ チ入出力ポート間でスループットのバランスが取れない

[0024] 以上より、スイッチSW1において選度変 でも、チャネルD01、D02、D03、D04でのデ できる。本実施の形態のディスク装団におけるデータ転 送方式としては、ファイバチャネルやインフィニパンド 1、D12、D13、D14上のデータ転送速度が低速 タDKAとディスクアレイDA間のスループットを向上 一夕転送速度は高速にできる。 つまり、ディスクアダプ 数と多項化、逆多国化が行われるので、チャネルロ1

[0025] 図6は、第1の実施の形態のディスク装置 において、ディスクドライブの増設方法を示した図であ 能性がある。そこでスイッチSW1では、ディスクアダ プタDKAとの間に、新規チャネルD05を増設してい 図6では図1に対して、ディスクグループR5とR め、スイッチSW1とSW2としてボート数の多いスイ ディスクアダプタ倒とのスループットバランスが協れる る。また、スイッチSW2の場合は筋規チャネルを増設 せず、チャネルD02の信号伝送速度を増加させること ル8本で接破し、ディスクアダプタとスイッチ間を4G ル8本で接続し、ディスクアダプタとスイッチ間を10 Gbpsのチャネル1本で接続する。このように、本実 ので、スイッチの速度変換機能が有効に働かなくなる可 は、スイッチとディスクアレイ間を1Gbpsのチャネ は、スイッチとディスクアレイ間を1Gbpsのチャネ 箱の形態のディスク装置は、スイッチのボート数に広じ で、ディスクアダプタ館とディスクアレイ館のスループ て、ディスクドライブを増設可能である。 このディスク ドライブ増設方法は、1ポート当たりに接続できるドラ イブ数が少ないATA (AT Attachment) ッチを使用している。ディスクドライブを増設すると、 スイッチのディスクアレイ館のスループットが勧笞し、 ットパランスを取っている。例えばスイッチSW1で 6が増設されている。ディスクドライブを増設するた bpsのチャネル2本で核板する。スイッチSW2で 方式ディスクドライブを増設するのに適用できる。

しと呼ぶ)がある。

ィスク装団は、ディスクコントローラDKCと、4個の F1X97V1DA1, DA2, DA3, DA4#54 る。ディスクコントローラDKCは、チャネルアダプタ CHA、キャッシュメモリCM、ディスクアダプタDK Aからなる。ディスクアレイDA1とディスクアダプタ DKAは、チャネルD01とスイッチSW1を介して接 睨する。同様に、ディスクアレイDA2はチャネルD0 2とスイッチSW2を介して、ディスクアレイDA3は チャネルロ03とスイッチSW3を介して、ディスクア レイDA4はチャネルD04とスイッチSW4を介し

て、それぞれディスクアダプタDKAと接破する。スイ ッチSW1、SW2、SW3とSW4は、第1の独筋の 形臨と同様に速度変換と多皿化、逆多田化を行うスイッ チとして機能する。本実施の形態におけるディスクアダ JADKAŁ, Z1"FSW1, SW2, SW3, SW 4*E.* <del>7</del>4*A*97*V*4*D*A1, *D*A2, *D*A3, *D*A4 との間のデータ転送方式は、ファイパチャネルを使用し Tいる. スイッチSW1, SW2, SW3, SW4はフ ァイバチャネルスイッチである。

[0027] 本政施の形臨におけるディスクアレイの格 成を、ディスクアレイDA1を倒に近べる。ディスクア と、D12上に接続した4個のディスクからなるディス クアレイと、D13上に依扱した4個のディスクからな るディスクアレイと、D14上に按板した4個のディス を例にとると、ディスクドライブDK1、DK2、DK イバチャネルアーピトレイテッドループ (以下FC-A りからなるディスクアレイ、からなる。チャネルロ11 のように、多数のドライブを一つのチャネル上に依成し レイDA1、DA2、DA3、DA4は、 同談のドライ てディスクドライブにアクセスする方法としては、ファ ブ梅成である。ディスクアレイDA1は、チャネルD1 1 上に接続した 4 個のディスクからなるディスクアレイ 3、DK4が、チャネルD11上に接続されている。 52 8

クドライプDK1、DK2、DK3、DK4の核粧形盤 を例として示す。各ディスクドライブの入出力ポートお よびスイッチSW1の入出力ポートは、送信機Txと受 [0028] 図10に、FC-ALの核殻形盤をディス ある。各ドライブの入出力ポートはファイバチャネルの 個機R×を有する。FC-ALの接続形態は、例えば図 10に示すように、各ドライブの入出力ポートおよびス イッチの入出力ポートをループ状に接続するトポロジで NI (Node Loop) ポートとして協信する。N しポートとは、ループ動作をする装団(ここではディス クドライブ) のポートである。スイッチSW1のディス **クアレイDA1接模倒入出力ポートは、ファイパチャネ** する。FLポートとは、FC-ALを接続可信なスイッ ルのFL(Fabric Loop) ポートとして機能 チのポートである。FLポートを有するループは、ファ 35

S スクアレイ部分の構成方法が異なる。本実施の形態のデ [0026] 図7に本発明の、第2の実施の形態である ディスク装団の構成を示す。本実施の形態のディスク装 **暨は、第1の実施の形態のディスク装置に対して、ディ** 

イバチャネルのパブリックループとして破悩するので、

## ディスクアダプタとディスクアレイをスイッチをかして接続したディスク装団

1. DK2, DK3, DK4H, Z1"FSW1ALU プとなる。パブリックループとは、ルーブ上のディスク ドライブが、スイッチを介してルーブ外のポートと道信 チャネルD11が形成するFC-ALはパブリックルー 可能なループである。よって、ディスクドライプDK

チャネルD01を介してディスクアダプタDKAと通信 可値である。以上、チャネルD11の技機形態を例に説 明したが、チャネルD12、D13、D14でも同様で ある。本英筋の形態のディスク模置においてRAIDシ 本実施の形倣では、4個のディスクドライブでRAID グルーブを構成しているが、RAIDグループを構成す 2、R3、R4を、それぞれRAIDグループとする。 ステムを構築する場合は、ディスクグループR1、R るドライブ数を4個に限るものではない。

ALの仕様から、チャネルD11, D12, D13, D 14上には、それぞれ最大126台までのディスクドラ 1, D12, D13, D14において、それぞれFC-ALを用いてディスクドライブを接続している。FCー を、チャネルD11, D12, D13, D14の媒体と [0029] 本実施の形態においては、チャネルロ1 イブが接続可値である。また、チャネルD01、D0 2、D03、D04の媒体として光ファイパケーブル してメタルケーブルを用いる。

ドライブ台数が均加できる。つまり、ディスク装置当り スク数団においては、ディスグドライブをFC-ALで 傍栊しているので、スイッチのポート当りに接続できる の記憶容量を増加させる効果がある。また、ディスクド ライブをメタルケーブルで接破することにより、ディス クドライブ年に高価な光インターフェイスを被償する必 **取がなくなるので、ディスクドライブのコストを下げる** [0030]以上説明したように、本実施の形態のディ

団は、ディスクコントローラとスイッチを二重化した点 [0031] 図8に本発明の、第3の実施の形態である ディスク藝団の構成を示す。本実施の形態のディスク藝 に怜徴がある。 本実施の形態において、ディスクアダブ タDKA1, DKA2と、スイッチSW1、SW2と、 ディスクアレイDA1との回のデータ転送方式は、ファ

ルアダプタCHA1と、キャッシュメモリCM1と、デ スイッチSW1、SW2と、ディスクアレイDA1から て做値する。ディスクコントローラDKC1は、チャネ **ラDKC2は、チャネルアダプタCHA2と、キャッシ** なる。スイッチSW1とSW2は、第1の実施の形態と 同様に斑斑蛇紋と多皿化、逆多皿化を行うスピッチとし イパチャネルを使用している。 本契筋の形態のディスク ィスクアダプタDKA1からなる。ディスクコントロー 被配は、ディスクコントローラDKC1、DKC2と、 ュメモリCM2と、ディスクアダプタDKA2からな

ッチ SW 2 をチャネルD 2 d で接続し、ディスクアダプ し、ディスクアダプタDKA2とスイッチSW1をチャ タDKA1 とスイッチSW2をチャネルD1 bで接続 ネルD2bで接続する。

た4個のディスクからなるディスクアレイと、D12と るディスクアレイと、D14とD24に接続した4個の ディスクからなるディスクアレイ、からなる。チャネル D11, D12, D13, D14, D21, D22, D [0032] ディスクアレイDA1を構成するディスク ネルD11およびD21の両チャネルと接続する。ディ スクアレイDA1は、チャネルD11とD21に接続し D22に接続した4個のディスクからなるディスクアレ イと、Dj3とD23に接続した4個のディスクからな 23、D24は、FC-ALでディスクドライブを接続 ドライブは、入出力ポートを2個有する。例えば、ディ 30 F 5 1 7 DK 1, DK 2, DK 3, DK 4 は, チャ

3、DK4の接続形態を例として示す。各ディスクドラ ドライブの入出力ポートおよびスイッチSW1、SW2 [0033] 図11に本実施の形態におけるFC-AL の接続形態を、ディスクドライブDK1、DK2、DK イブは、それぞれNLポートを2個有する。各ディスク の入出力ポートは、送信機T×と受信機R×を有する。

入出力ポートは、FLポートである。チャネルD11に 2、DK3、DK4をループ状に接続する。同様にチャ スイッチSW1、SW2のディスクアレイDA1接続側 より、スイッチSW1、ディスクドライプDK1、DK ネルD21により、スイッチSW2、ディスクドライブ DK1, DK2, DK3, DK4をループ状に接続す

る。これら2個のループは、ファイバチャネルのパブリ てディスクアダプタDKA1またはDKA2と通信可能 である。以上、チャネルD11、D21の接続形態を例 DK3、DK4は、スイッチSW1またはSW2を介し ックループであり、ディスクドライブDK1、DK2、

に説明したが、チャネルD12、D13、D14、D2 2、D23、D24でも同様である。本実施の形態のデ は、ディスクグループR1、R2、R3、R4を、それ ぞれRAIDグループとする。本実施の形態では、4個 のディスクドライブでRAIDグループを構成している が、RAIDグループを構成するドライブ数を4個に関 イスク装置においてRAIDシステムを構築する場合 るものではない。

イブは、ディスクアダプタDKA1およびDKA2のど クアレイDA1にアクセスできる。逆に、スイッチSW [0034] ディスクアレイDA1内の全ディスクドラ ちらからでもアクセス可能である。本実施の形態のディ 1、SW2故障時の迂回経路として使用する。例えばス イッチSW1が故障した場合でも、ディスクアダプタD KA1はチャネルD1bとスイッチSW2框由でディス スク装置は、チャネルD1b、D2bをスイッチSW 20

ネルD 2 b とスイッチ S W 1 経由でディスクアレイDA 2が故障した場合は、ディスクアダプタDKA2はチャ 1 にアクセスできるので、信頼性が高いディスク装置が [0035] 図12に本発明の、第4の実施の形態であ 5ディスク装団の構成を示す。本実施の形態のディスク ·ッチSW1、SW2間を接続するチャネルD3a、D F1207879DKA1, DKA22, Z1"FSW アレイDA1からなる。スイッチSW1とSW2は、第 ディスクアダプタDKA1とスイッチSW2をチャネル 装置は、第3の実施の形態のディスク装置に対して、ス 1、SW2と、ディスクアレイDA1との間のデータ転 送方式は、ファイバチャネルを使用している。本実施の 1の実施の形態と同様に速度変換と多重化、逆多重化を **行うスイッチとして機能する。ディスクコントローラD** イスクコントローラDKC2は、チャネルアダプタCH モリCM1と、ディスクアダプタDKA1からなる。デ 1. DKC2Ł, スイッチSW1, SW2Ł, ディスク KC1は、チャネルアダプタCHA1と、キャッシュメ A2と、キャッシュメモリCM2と、ディスクアダプタ DKA2からなる。 ディスクアダプタDKA1とスイッ チSW1をチャネルD1aで接続し、ディスクアタブタ 3 bを散けた点に特徴が有る。本実施の形態において、 形態のディスク装置は、ディスクコントローラDKC DKA2とスイッチSW2をチャネルD2aで接続し、

D1bで接続し、ディスクアダプタDKA2とスイッチ SW1をチャネルD2bで撩舵する。さらに、スイッチ D22に接続した4個のディスクからなるディスクアレ るディスクアレイと、D14とD24に接続した4個の ドライブは、入出力ポートを2個有する。例えば、ディ D11, D12, D13, D14, D21, D22, D スクドライブを接続する。ディスクアレイDA1内の金 ディスクドライブは、ディスクアダプタDKA1および DKA2のどちらからでもアクセス可能である。本実筋 する場合は、ディスクグループR1、R2、R3、R4 [0036] ディスクアレイDA1を構成するディスク スクドライプDK1, DK2, DK3, DK4は, チャ ネルD11およびD21の両チャネルと接続する。ディ スクアレイDA1は、チャネルD11とD21に接続し た4個のディスクからなるディスクアレイと、D12と イと、D13とD23に接続した4個のディスクからな ディスクからなるディスクアレイ、からなる。 チャネル 23. D24は、図11に示すようにFC-ALでディ の形態のディスク装置においてRAIDシステムを構築 を、それぞれRAIDグループとする。本実施の形態で は、4個のディスクドライブでRAIDグループを構成 しているが、RAIDグループを構成するドライブ数を SW1とSW2をチャネルD3a、D3bで接続する。

50 チャネルD2 a上のスループットに何限され、2Gbp [0037] ディスクアダプタDKA1, DKA2とデ

ィスクアダプタDKA1は、チャネルD1aとSW1を 介してディスクアレイDA1にアクセスする経路 (経路 1) と、チャネルロ1 bとスイッチ SW2 とチャネルロ 3 a とスイッチ S W 1 を介してディスクアレイDA 1 に アクセスする経路(軽路2)を有する。同様に、ディス クアダプタDKA2は、チャネルD2aとSW2を介し てディスクアレイDA1にアクセスする経路·(経路3) **常時(スイッチ故障無しの場合)について説明する。** イスクアレイDA1のアクセス経路について、先ず、 8

S

വ

**特開2003-3030** 

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

セスする経路 (経路4)を有する。一方、スイッチ故邸 と、チャネルD2bとスイッチSW1とチャネルD3b とスイッチSW2を介してディスクアレイDA1にアク 時は、チャネルD1b、D2bを迂回経路として使用す る。例えばスイッチSW1が牧陣した場合でも、ディス クアダプタDKA1はチャネルD16とスイッチSW2 スイッチSW2が故障した場合は、ディスクアダプタD KA2はチャネルD2bとスイッチSW1種由でディス 紙由でディスクアレイDA1にアクセスできる。逆に、 クアレイDA1にアクセスできる。 【0038】次に、本実施の形倣におけるディスクアダ b、D3a、D3b上のデータ伝送遊取をチャネル当り 2Gbps 2 C. Ft \* ND11, D12, D13, D 14、D21、D22、D23、D24上のデータ伝送 き、スイッチSW1とディスクアレイDA1回の総スル ブターディスクアレイ間のスルーブットについて説明す る。例として、チャネルD1a、D1b、D2a、D2 速度をチャネル当り1Gbpsであるとする。このと

ープットは4Gbpsである。ディスクアダプタDKA クセスすることにより、総スループットは4Gbpsと と、ディスクア・レイDA1億のスループットが共に4G 1とスイッチSW1間は、上記経路1および経路2でア **ちpsであるので、ディスクアダプタDKA1とディス** なる。スイッチSW1のディスクアダプタDKA1伽 クアレイDA1間のスループットは4Gbpsとなる。

同様に、スイッチSW2とディスクアレイDA1間の総 スルーブットは4Gbpsである。ディスクアダプタD でアクセスすることにより、総スループットは4Gbp ディスクアダプタDKA1とディスクアレイDA1間の KA2とスイッチSW2同は、上記額路3および矩路4 sとなる。スイッチSW2のディスクアダプタDKA2 **歐と、ディスクアレイDA1億のスループットが共に4** スクアレイDA2回のスルーブットは4Gbpsとなる 第3の実施の形態 (図8) において、上記のチャネル当 **りスループット値を適用すると、チャネルD1b、D2** Gbpsであるので、ディスクアダプタDKA2とディ スループットは、チャネルD18上のスループットに飼 限され、2Gbpsとなる。同様に、ディスクアダプタ **しをスイッチ故障時の迂回経路としか使用しないので、** DKA2とディスクアレイDA1肌のスループットは、

ネルD 1 a で接続し、ディスクアダプタDKA2とスイ

る。ディスクアダプタDKA1とスイッチSW1をチャ

ら、本史筋の形態によれば、ディスクアダプタースイッ ターディスクアレイ間スループットを4Gbpsにする クアダプターディスクアレイ間の総スループットが高い 8となる。第3の実施の形態において、ディスクアダプ ためには、チャネルD1点およびD28のデータ伝送渡 度を、それぞれ4Gbp8に高める必要がある。以上か **チ間のチャネル当りデータ伝送速度が低くても、ディス** ディスク岐四が奥鳴できる。 (0039)

Fの効果がある。ディスクアダプタとディスクアレイ関 ク较同を投供できる。また、信頼性の高いディスクアレ 「兜明の効果」以上説明したように、本発明によれば以 た、ディスクアダプタとディスクアレイ間のスルーブッ イを有するディスク装置を提供できる。また、借頼性が レイ間のスループットが高く低コストなディスク装置を 商いディスクアダプタとディスクアレイ闘ネットワーク を有するディスク藝囮を提供できる。また、俗類性およ びスルーブットが高いディスクアダプタとディスクアレ できる。また、高スルーブットを維持できるディスク藝 トが高く、且つディスクドライブ接続台数が多いディス 5. また、ディスクからの脱み出しおよびディスクへの **昏さこみを高スループット化できるディスク装配を提供** 回を抱供できる。また、ディスクアダプタとディスクア のスループットが高いディスク数固を提供できる。虫 イ間ネットワーク、を有するディスク数固を提供でき お供できる。

【図1】本発明の約1の実施の形態のディスク装置を示 【図面の簡単な説明】 す図である。

[図2] 本発明に用いるスイッチの構成を示す図であ 【図3】 本発明に用いるスイッチの構成を示す図であ 【図4】本発明に用いるスイッチの動作を示す図であ

【図5】 本発明に用いるスイッチの動作を示す図であ

[図6] 本発明第1の実施の形態に対して、ディスクド ライブを増散する方法を示す図である。

【図7】本発明の第2の実施の形態のディスク装置を示 す図である。 【図8】本発明の第3の実施の形態のディスク装置を示 「図である。

[図10] FC-ALによる接税形態を説明する図であ [図9] 従来のディスク装置を示す図である。

【図11】FC-ALによる接続形態を説明する図であ

【図12】本発明の第4の実施の形態のディスク装置を 示す図である。

[符号の説明]

DKC, DKC1, DKC2 74A9コントローラ CHA、CHA1、CHA2 チャネルアダプタ

DKA, DKA1, DKA2 74ADTY79 CM, CM1, CM2 キャッシュメモリ DA. DA1~DA4, F1X0FV1 DK1~DK4 ディスクドライブ R1~R6 ディスクグループ

C1~C4, D1~D4, D01~D05, D11~D 14. D21~D24. D1a. Dib. D2a. D2

D. D38. D3b F+\*/ P1~P5 スイッチボート SW1~SW4 A1"F

·out1~out5 クロスパスイッチ出力ポート in1~in5 クロスパスイッチ入力ポート **CTL スイッチコントローラ** 25 XSW クロスパスイッチ

ENC1、ENC2 8B10B変換エンコーダ DEC1, DEC2 8B10B変換デコーダ パラレルシリアル変換装置 SP1、SP2 シリアルバラレル変換装個 BM1, BM2 パッファメモリ T1, T2, T3 PS1, PS2

T4、T5、T6 フレームの時回 T× 说信题

Rx 受信機

NL NL X-F

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

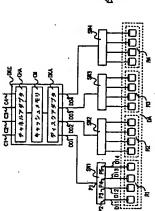
S

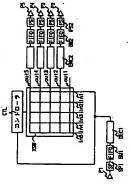
特開2003-30305

(図1)

[図2]

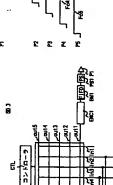
8





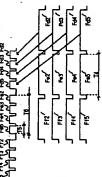
[図4]

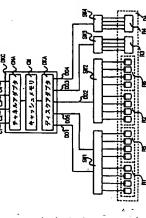
(⊠3)



[885]

(図()





- 11 -

特開2003-303055

ディスクアダプタとディスクアレイをスイッチ を介して接続したディスク装置

**特開2003-30305** 

